RESEARCH

UDOT

Keeping Utah Moving

Utah Department of Transportation - Research Division
4501 South 2700 West - P.O. Box 148410 - SLC,UT 84114-8410

Report No. UT-19.12

A DATA FUSION APPROACH FOR EXTRACTING HIGHWAY MAINTENANCE FEATURES

Prepared For:

Utah Department of Transportation
Research Division

Submitted By:

Utah State University
Department of Civil and Environmental Engineering
Department of Computer Science

Authored By:

Ziqi Song, Ph.D.
Mohammadreza Javanmardi
Xiaojun Qi, Ph.D.

Final Report
June 2019

DISCLAIMER

The authors alone are responsible for the preparation and accuracy of the information, data, analysis, discussions, recommendations, and conclusions presented herein. The contents do not necessarily reflect the views, opinions, endorsements, or policies of the Utah Department of Transportation or the U.S. Department of Transportation. The Utah Department of Transportation makes no representation or warranty of any kind, and assumes no liability therefore.

ACKNOWLEDGMENTS

The authors acknowledge the Utah Department of Transportation (UDOT) for funding this research, and the following individuals from UDOT on the Technical Advisory Committee for helping to guide the research:

- Tim Ularich
- Ryan Ferrin
- Kevin Griffin
- Vincent Liu

TECHNICAL REPORT ABSTRACT

1. Report No. UT-19.12	2. Government Accession No. N/A	3. Recipient's Catalog No. N/A
4. Title and Subtitle A Data Fusion Approach for Extracting Highway Maintenance Features		5. Report Date June 2019 6. Performing Organization Code
7. Author(s) Ziqi Song, Mohammadreza Javanmardi, Xiaojun Qi		8. Performing Organization Report No.
9. Performing Organization Name and Address Utah State University Department of Civil and Environmental Engineering Department of Computer Science Logan, UT 84322		10. Work Unit No. 5H08039H 11. Contract or Grant No. 18-8440
12. Sponsoring Agency Name and Address Utah Department of Transportation 4501 South 2700 West P.O. Box 148410 Salt Lake City, UT 84114-8410		13. Type of Report & Period Covered Final Report January 2018 to June 2019 14. Sponsoring Agency Code PIC No. UT16.205

15. Supplementary Notes

Prepared in cooperation with the Utah Department of Transportation and the U.S. Department of Transportation, Federal Highway Administration

16. Abstract

Most existing studies investigating highway feature extraction in the literature focus on a single source of data, e.g., aerial LiDAR or imagery data. Because each source of input data contains its own characteristic visual features, the fusion of the characteristic and complementary features acquired from both sensors can provide more reliable information to facilitate the accurate detection of maintenance features and therefore improve the robustness of the algorithm, especially when data from either of the input sources are of low quality. In the first stage of this project, we propose an effective method to automatically detect traffic signs and light poles from mobile LiDAR data in the I-15 highway without any preprocessing or learning steps. A set of extensive experiments have been carried out on the data sets, which are captured by UDOT along the I-15 highway. The results demonstrate the robustness of the proposed method in detecting almost all traffic signs and light poles. In the second stage of this project, we propose a novel deep learning method to accurately detect traffic signs by fusing the characteristic and complementary features in automatically registered airborne geo-referenced images and airborne LiDAR data. A set of extensive experiments has been carried out on the airborne geo-referenced images and airborne LiDAR data captured by USU along the I-15 highway. The results demonstrate the robustness of the proposed method in detecting traffic signs.

17. Key Words Traffic Sign Detection, Convex Optimization, Deep Learning, Convolutional Neural Network		18. Distribution Statement Not restricted. Available through: UDOT Research Division 4501 South 2700 West P.O. Box 148410 Salt Lake City, UT 84114-8410 www.udot.utah.gov/go/research		23. Registrant's Seal N/A
19. Security Classification (of this report) Unclassified	20. Security Classification (of this page) Unclassified	21. No. of Pages 46	22. Price N/A	

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ACRONYMS

LiDAR	Light Detection and Ranging
CNN	Convolutional Neural Network
HOG	Histogram of Oriented Gradient
SVM	Support Vector Machine
MLS	Mobile Light detection and ranging Scanning
PCA	Principal Component Analysis
LDH	Local Descriptor Histogram

EXECUTIVE SUMMARY

Departments of transportation (DOTs) primarily rely on manual data collection and more recent mobile LiDAR (light detection and ranging) data collection methods, which are often time-consuming and costly, to update their maintenance feature inventory. Most existing studies investigating highway feature extraction in the literature focus on a single source of data, e.g., aerial LiDAR or imagery data. Because each source of input data contains its own characteristic visual features, the fusion of the characteristic and complementary features acquired from both sensors can provide more reliable information to facilitate the accurate detection of maintenance features and therefore improve the robustness of the algorithm, especially when data from either of the input sources are of low quality. In this study, we focus on selected maintenance features, i.e., traffic signs and light poles, for algorithm development purposes. That being said, the algorithms and workflows developed by this study can be readily extended to other maintenance features without difficulty.

In the first stage of this project, we propose an effective method to automatically detect traffic signs and light poles from mobile LiDAR data on the I-15 highway without any preprocessing or learning steps. Specifically, we first use the surface reconstruction algorithm to extract the normal vectors of the points as one of the characteristic features and apply *k*-means on the characteristic features of the points to automatically segment the data into road or nonroad points. We then employ sliding cuboids to search for high-elevated objects that are located near the borders and on top of the road points. We further employ the random sample consensus algorithm to remove outliers and keep the points that fall on the perpendicular planes to the road trajectory. Finally, we introduce a modified seeded region growing algorithm to remove noisy points and incorporate the shape information to reject the false objects. Extensive experiments have been carried out on the data sets, which are captured by UDOT along the I-15 highway. The results demonstrate the robustness of the proposed method in detecting almost all traffic signs and light poles.

In the second stage of this project, we propose a novel deep learning method to accurately detect traffic signs by fusing the characteristic and complementary features in automatically registered airborne geo-referenced images and airborne LiDAR data. Specifically, we first

segment the airborne RGB images to road and nonroad segments. Second, we geo-reference the corresponding LiDAR data and find the road sections in this airborne LiDAR data. Third, we use the height information of the road points in LiDAR data to extract high elevated objects above the road. Fourth, we segment the extracted objects to different regions (traffic sign candidates) using Euclidean distance-based clustering. Fifth, we find the corresponding traffic sign candidates in RGB images. Finally, we extract convolutional neural network (CNN) features of traffic sign candidates and represent them in a convex optimization framework to classify them as traffic sign or nontraffic sign classes. A set of extensive experiments has been carried out on the airborne geo-referenced images and airborne LiDAR data captured by USU along the I-15 highway. The results demonstrate the robustness of the proposed method in detecting traffic signs. Specifically, we provide the true positives, false positives, and true negatives of the proposed method on all nine sections of the data set. In total, there are 17 traffic signs in these sections. The proposed method extracts 24 traffic sign candidates. In the classification process, 14 out of 24 candidates are correctly classified as traffic signs (true positives), three out of 24 candidates are incorrectly classified as traffic signs (false positives), and seven out of 24 candidates are correctly classified as nontraffic signs (true negatives). In other words, the proposed method is able to successfully extract 14 out of 17 traffic signs and achieve detection accuracy of 82.35%.

1.0 INTRODUCTION

1.1 Problem Statement

Departments of transportation (DOTs) primarily rely on manual data collection and more recent mobile LiDAR (light detection and ranging) data collection methods, which are often time-consuming and costly, to update their maintenance feature inventory. In a recent study, the research team at Utah State University (USU) investigated whether aerial LiDAR data can be a more cost-effective means than the mobile counterparts to identify highway features in areas where large-scale changes may have occurred and a way to keep the feature inventory current [1-3]. In the airborne data collection conducted by USU in 2015, both aerial LiDAR point cloud and high-resolution aerial imagery data were obtained. Results of the study demonstrate that aerial LiDAR is a promising technology in detecting some highway features, such as guardrails, medians, and light poles as well as large road signs. It is also found that geo-referenced high-resolution aerial imagery data can be used as an alternative means to identify highway features. Moreover, aerial imagery data are particularly suitable for developing automated procedures for highway feature extraction, which offers great potential in time and cost savings when updating feature inventory. With the advent of unmanned aerial vehicle (UAV) technology, high-resolution aerial images will be much more affordable and easily accessible for transportation agencies in the future. Although the current multispectral aerial image data set was collected with a fixed-wing plane, the methodology developed for the current data set could be readily transferable to any UAV-based data collection platform. However, aerial imagery data do not contain any elevation information, which is critical for identifying some highway features.

Most existing studies investigating highway feature extraction in the literature focus on a single source of data, e.g., aerial LiDAR or imagery data. Based on our previous studies, we found that each data collection method has advantages and limitations. The most effective approach to achieve the maximum level of accuracy and completeness is to combine data collected from multiple sources. The data set obtained from our recent UDOT study contains both aerial LiDAR and aerial imagery data and further provides us with a unique opportunity to pursue this research direction. Therefore, it is imperative to develop a data fusion approach that utilizes both aerial LiDAR and aerial imagery data to compensate for the limitations of both

methods. In addition, automated feature extraction using LiDAR data is an important aspect of large-scale applications. This study also aims to develop an efficient workflow for extraction of highway maintenance features using LiDAR data.

In this study, we focus on selected maintenance features, i.e., traffic signs and light poles, for algorithm development purposes. That being said, the algorithms and workflows developed by this study can be readily extended to other maintenance features without difficulty. We choose to focus on traffic signs because they play an important role in any transportation system, e.g., they notify drivers about the current road situation, warn drivers to drive at the prescribed speed, and provide drivers with useful information for safe driving. Therefore, detection and identification of traffic signs is an important research direction, which is of great significance to prevent road traffic accidents and reinforce driver safety. In addition, reliable traffic sign detection and identification lead to better autonomous driving systems, more efficient road sign inspection, and faster transportation. Light poles are also investigated because they are not covered by the current UDOT inventory survey.

1.2 Objectives

The previous work mainly focuses on detecting maintenance features in images captured by on-board cameras or in mobile light detection and ranging (LiDAR) data captured by on-board laser scanners. Because each source of input data contains its own characteristic visual features, the fusion of the characteristic and complementary features acquired from both data sources can provide more reliable information to facilitate the accurate detection of maintenance features and therefore improve the robustness of the algorithm, especially when data from either of the input sources are of low quality. In this project, we propose a novel deep learning method to accurately detect traffic signs by fusing the characteristic and complementary features in automatically registered airborne geo-referenced images and airborne LiDAR data.

1.3 Scope

This project consists of two major stages:

- In the first stage, we propose an effective method to automatically detect traffic signs and light poles from mobile LiDAR data of the I-15 highway without any preprocessing or learning steps. This method consists of three major steps: road points extraction, traffic sign and light pole detection and classification, and postprocessing for false objects rejection.

- In the second stage, we design and implement a method that utilizes the complementary information captured from airborne RGB images and airborne LiDAR data to detect traffic signs in highway areas. This method consists of three major components: road extraction, traffic sign candidate detection, and traffic sign classification.

1.4 Outline of Report

Chapters included in this report are as follows:

- Introduction
    - Problem Statement
    - Objectives
    - Scope
- Literature Review of Previous Studies
- Data Collection
- Traffic Sign and Light Pole Detection in Mobile LiDAR
    - Road Point Extraction
    - Traffic Sign and Light Pole Extraction and Classification
    - Postprocessing
    - Data Evaluation
        - Qualitative Results
        - Quantitative Results
- Traffic Sign Detection Using Registered Geo-Referenced RGB Images and Airborne LiDAR data
    - Road Extraction
    - Traffic Sign Candidate Detection
    - Traffic Sign Classification
    - Data Evaluation
        - Qualitative Evaluation
        - Quantitative Evaluation

- Conclusions
    - Overview
    - Conclusions
    - Findings
    - Challenges

2.0 DATA COLLECTION

The mobile LiDAR data for the first stage of this project were collected by a UDOT (Mandli) vehicle with the following equipment: Velodyne LiDAR sensor, laser road imaging system, laser rut and crack measurement system, road surface profiler, and position orientation system.

For the second stage of the project, we use geo-referenced RGB images and airborne LiDAR data, which were simultaneously captured from different sections of the I-15 highway located in Utah, United States, to more accurately identify maintenance features than using either geo-referenced RGB images or airborne LiDAR data. To collect the data, USU installed a camera and a LiDAR device on an airplane, which flew above the I-15 highway.

The USU airborne LiDAR system is mounted in a single-engine Cessna TP206 aircraft (Figure 2.1). The system consists of a LiDAR scanner, IMU, and flight navigation unit. The LiDAR instrument consists of a Riegl Q560 transceiver and Novatel SPAN LN-200 GPS/IMU positioning and orientations system. Depending on the flight height, the LiDAR scanner is able to collect data at a pulse rate of 250,000 shots/seconds. Together with the LiDAR system, the USU airborne system is also equipped with multispectral and thermal infrared cameras, which can be used for aerial photos. The camera system is composed of four ImperX 4820 Monochrome cameras with 4872 x 3248 pixels per camera. The cameras are also equipped with interface filters in the blue, green, red, and near-infrared (NIR) centered at 0.472, 0.562, 0.655, and 0.80 μm, respectively.

Figure 2.1 USU Cessna TP206 Research Aircraft

The input geo-referenced RGB images and the airborne LiDAR data contain information about the road, buildings, parking lots, vegetation, traffic signs, billboards, bridges, etc. In total, we are provided with 11 mosaic maps (produced by using the captured RGB images) and 14 airborne LiDAR data sets. In our method, we concatenate the 14 LiDAR data sets to obtain a mosaic LiDAR map to facilitate the registration of RGB geo-referenced images and the LiDAR data.

3.0 LITERATURE REVIEW

3.1 Overview

In this chapter, we introduce some of the existing traffic sign detection methods and our proposed framework, which detect traffic signs in the highway areas by using registered geo-referenced images and airborne LiDAR data. Specifically, in Section 3.2, we provide the literature review of representative traffic sign detection methods and the type of data sets utilized in each research work.

3.2 Literature Review of Traffic Sign Detection

Traffic signs are considered one of the most important maintenance features, as they inform drivers or recent autonomous cars on current road conditions. Various mature traffic sign extraction methods have been proposed using different types of data captured by different types of sensors. Two common categories of these methods utilize images and light detection and ranging scanning (LiDAR) data. Here, we briefly review several representative approaches that use camera images and LiDAR data.

Substantial image-based algorithms have been proposed to detect and recognize traffic signs. For example, Soheilian et al. [4] present an automatic approach toward utilizing color information to identify the silhouette of signs in every individual image. The authors then propose a multiview constrained 3D reconstruction algorithm to provide an optimum 3D silhouette to detect traffic signs. Adam and Ioannidis [5] propose to extract traffic signs by using color images acquired by a camera mounted on a moving vehicle. They detect the regions of interest (ROIs) and classify them as traffic or nontraffic signs by feeding the regions' histogram of oriented gradient (HOG) descriptors to a trained support vector machine (SVM). Khalid et al. [6] estimate a global threshold value using the correlation property of a given image and segment the regions of traffic signs based on the global threshold and morphological operations. They further detect traffic signs by feeding HOG descriptors to a trained SVM-*k*-nearest neighbor classifier. Despite the favorable performance of these algorithms using camera images, visual

features of traffic signs such as color, shape, and appearance are often sensitive to illumination conditions, angles of view, etc.

Recently, researchers have proposed various methods utilizing LiDAR technology for traffic sign detection. However, the number of published works in this aspect is relatively small. Most of these methods use mobile light detection and ranging scanning (MLS) data, as they usually have better quality and density than airborne LiDAR. For instance, Pu et al. [7] introduce one of the pioneer studies in detecting and distinguishing traffic signs using LiDAR data. In this work, they initially segment data into one of the three coarse categories, including the ground surface, the objects on the ground, and the objects off the ground. They further use the size, shape, and orientation information to classify the on-ground points to more detailed classes such as traffic signs. Yokoyama et al. [8] employ principal component analysis (PCA) to extract pole-like objects from MLS data and classify them into utility poles, lamp posts, or street signs. Yu and Li [9] propose a voxel-based upward growing method to remove the ground points and a voxel-based normalized cut to segment the remaining MLS point clouds data into street light poles, traffic signposts, or bus stations. Riveiro et al. [10] employ the geometric and radiometric information of retro-reflective traffic signs in the segmentation process to compute the optimal intensity threshold to separate traffic signs from backgrounds. Lehtomaki et al. [11] propose to remove ground and building points from the original data to reduce the search space. They then incorporate three sets of features, i.e., local descriptor histogram (LDH), spin images, and general shape and point distribution, to segment data into trees, lamp posts, traffic signs, cars, penetrations, and hoardings. The aforementioned method uses the three-dimensional information or reflectiveness of traffic signs for detection purposes. However, the quality of the three-dimensional information is relatively low in the airborne LiDAR data due to the different angles of view, a large number of outliers, and different object poses. This degrades efficiency of these algorithms when dealing with airborne LiDAR data.

4.0 TRAFFIC SIGN AND LIGHT POLE DETECTION IN MOBILE LIDAR

4.1 Framework of the Proposed Method

In this study, we propose an effective method to automatically detect traffic signs and light poles from MLS point clouds in highway areas without any preprocessing or learning steps. This method consists of three major steps: road points extraction, traffic sign and light pole detection and classification, and postprocessing for false objects rejection:

- **Road Extraction:** Extract the road in the following steps:
    - Using the surface reconstruction algorithm to extract normal vectors of the points as one of the selected characteristic features.
    - Proposing an unsupervised road point extraction scheme by applying the *k*-means clustering on the characteristic features.

- **Traffic Sign and Light Pole Candidate Detection and Classification:** Detect traffic sign and light pole candidates and classify them in the following steps:
    - Designing a sliding cuboid to identify groups of candidate points by searching for the high elevated objects above or beside the roads.
    - Employing the random sample consensus (RANSAC) algorithm in a novel and unique way to select the robust candidate points by removing high elevated outliers that do not represent perpendicular planes along the vehicle trajectory.
    - Using the height property to classify the candidates into traffic sign and light pole classes.

- **Postprocessing:** Remove the false objects in the following steps:
    - Proposing a LiDAR-modified seeded region growing algorithm to remove the noisy points around the objects.
    - Introducing a two-step postprocessing method to remove false positive objects.

4.1.1 Road Points Extraction

The original point clouds, as the input, contain a large number of four-dimensional points. Each point includes the global positional values (i.e., *x*, *y*, and *z*) and the intensity value. The intensity is a measure of the returned strength of the laser pulse that is generated from the point. To reduce the computational time and facilitate processing, we segment the original point clouds to a predefined number of sections. Because the elevation of the points may gradually vary along the uneven or hilly road direction, the points in a smaller section tend to have more similar elevation than the points in a larger section. Each section is a portion of the data set along the length of the road (i.e., *y* axis). We then process each section separately in the following steps, and the resultant points are concatenated to uniquely represent the road.

Normal Vector Extraction: This step aims to estimate the normal vector for each data point to represent its orientation. We adopt the surface construction method, which uses a fixed number of neighboring points to fit a local plane to determine the normal vector of each data point in a section. For each section, we include the location information (e.g., *x*, *y*, and *z*) of all the points in a set of three-dimensional data. We first find six nearest neighbors of a point in the set and fit a local plane to the neighborhood plane. We then compute the centroid of the points in the neighborhood plane. We finally employ the PCA method on the neighborhood plane to calculate the smallest principal vector as the normal vector.

Data Points Clustering: This step clusters the data points to road or nonroad classes by incorporating the normal vector of the points as one of the characteristic features. Specifically, for each data point in the point clouds, we propose to concatenate its three-dimensional normal vector (i.e., orientation) with its *z* value (i.e., elevation) and intensity value (i.e., reflectivity) to construct a five-dimensional feature vector to represent its road characteristics. We then employ the *k*-means clustering algorithm, which is a powerful unsupervised method, with *k* being 2, to group all the data points into either the road or nonroad cluster.

4.1.2 Traffic Sign and Light Pole Extraction and Classification

Traffic signs are mostly located near the border or on top of the roads to be visible for drivers. In addition, light poles are raised sources of light on the edges of a road or path. We use

this prior information to eliminate the data points that are located off the road and eliminate the off-road counterfeit objects such as billboards and buildings that might have similar characteristics such as intensity and elevation as the traffic signs or light poles. As a result, we effectively reduce the search space in high-density MLS point clouds and reduce the computational time. In addition, we utilize some observational statistics (e.g., height, elevation, and planar projection of traffic signs or light poles) to further remove the points that are unlikely to be road points.

Cuboid Searching: This step aims to utilize a sliding rectangular cuboid to find traffic signs located near the border or on top of the road and the light poles that are situated near the borders of the roads. Specifically, we use the road points in each section to find their ranges for the x and y values. We then select the points whose x and y values fall into these calculated ranges. It should be noted that the selected points may correspond to roads, traffic signs, light poles, bridges, moving vehicles, and other objects near or on the roads. To solve this issue, we design a rectangular cuboid search strategy to identify the points that correspond to traffic signs and light poles. We first define the starting point in the x–y plane of each section at the minimum x and y values of the points. We then put a rectangular cuboid with the dimension of 4×4×inf (i.e., no limitation in the z direction) at the starting point and continue moving this sliding cuboid along 4 m at the x direction or 4 m at the y direction until the sliding cuboids cover all the points. At each location of the cuboid, we have a set of points that can belong to traffic signs, light poles, and any other objects. We observe that a majority of traffic signs have an elevation range of more than 3 m and are located above 1.5 m from the road surface. Moreover, all the light poles have the range of elevation more than 3 m. Therefore, we first utilize the height information to filter out obvious outliers by removing a set whose range of z values is less than 3. For each kept set, we further remove any points with the height values less than an adaptive threshold, which is computed by a predefined value (e.g., 1.5) plus the mean height value of all the points in the set. The remaining points in a kept set form a candidate set for a traffic sign or a light pole.

Plane Fitting: This step focuses on utilizing the planar property to extract traffic signs and light poles by finding the points that represent planes. In order to identify the region of interest in a candidate set for each cuboid location, we adopt the random sample consensus

(RANSAC) algorithm to discard the points that do not belong to a plane estimated by a sufficient number of inliers. RANSAC is an iterative method for estimating parameters of a mathematical model from a set of data points containing outliers. The input to RANSAC contains a candidate set, a parametrized fitting model, and two confidence parameters (e.g., the maximum distance and the maximum angular distance). The parametrized model is a reference normal vector, whose element is the projected value at the *-y* direction. This vector is used as an orientation constraint to fit a plane that has an approximate normal vector similar to the reference vector. RANSAC fits a plane to the input points to achieve the maximum distance and the maximum angular distance from the inlier points to the plane.

Intensity Threshold: This step is to use highly reflective property to remove points that do not belong to traffic signs and light poles. Because traffic signs are covered by highly reflective materials to make them visible during inclement weather conditions, day or night, and light poles are manufactured with metals, their corresponding point clouds normally have higher intensities. We can then use the intensity information to extract traffic signs and light poles. In the proposed method, we discard a candidate set if the average intensity of the points within this set is less than a predefined threshold. All the remaining candidate sets for all sections are kept as candidates for traffic signs and light poles.

Classification: The height property is used to classify the data into traffic sign and light pole classes. Because traffic signs in highway areas usually have lower elevation (height) than light poles, we use this prior information to estimate a threshold for the height of the candidates to automatically classify each candidate as either a traffic sign or light pole. For each set, we first find the range of *z* of its points. Second, we use an estimated threshold to segment the candidate points into two groups: Group 1 consists of all candidate points whose *z* ranges are less than the estimated threshold; Group 2 consists of the remaining candidate points. Third, we calculate the average *z* ranges for the candidate points in Group 1 and Group 2 as μ_1 and μ_2. Fourth, we update the estimated threshold by means of μ_1 and μ_2. Fifth, we repeat the process by segmenting the candidate points into two groups using the new estimated threshold until the difference between the threshold values in two iterations is smaller than a predefined number.

4.1.3 Postprocessing

We aim to utilize shape information to eliminate the false objects from the classification results. To this end, we propose a two-step postprocessing method to achieve more reliable performance by removing the outliers of the projection points in the *x*–*z* plane and taking the shape of the cleaned projection points into consideration. We observe that the *x*–*z* projection of the points representing traffic signs and light poles is approximately rectangular shaped. The projection of the points that belong to false objects (e.g., trees, bridges, buildings, and vehicles on the road) usually does not exhibit a rectangular shape. Because outlier points can be randomly located around each object, they may make the overall convex hull of the projected points to be nonrectangular. To address this issue, we propose the modified seeded region growing algorithm to eliminate the outliers around the object. We then use the rectangular shape information to eliminate the false objects.

Modified Seeded Region Growing: We propose the modified seeded region growing algorithm on the LiDAR data to group the potential object points and eliminate the outlier points around a candidate object. The modified seeded region growing method is able to robustly identify the same potential object points and recognize the same outlier points regardless of the initially selected seeds. It is also able to quickly identify a different number of neighbors for each data point based on its cloud density. At a location of a candidate cuboid in each section, the points in the traffic sign candidate set may represent a traffic sign object or a false object. Similarly, the points in a light pole candidate set may represent a light pole object or a false object. Regardless of the object types, there might be some outlier points around the objects. To eliminate these outliers, we randomly select a seed point and grow a uniform connected region from this seed. A point that has not been assigned to any other region is added to the seeded region if and only if the point is in the neighbor of the region and its distance to the region is less than a predefined threshold. Otherwise, the point will be a new seed for another region.

To achieve this goal, for each traffic sign candidate set or light pole candidate set, we calculate its *x*–*z* projection. Any point in the *x*–*z* plane is selected as the seed; then, its six nearest neighbors, which can fit a local plane, are selected. A neighbor is added to the region containing the seed if its Euclidean distance to the centroid of the region is less than 0.5 m and it has not

been assigned to any region. The algorithm continues to find the six nearest neighbors of each of the added points and repeats the same process to grow the region. When no neighbor can be added to any existing region, the algorithm finds a point that has not been assigned to any region as the seed and starts repeating the same growth process. We save all the points identified in the growing regions in a region list.

False Object Removal: We propose to use the shape information in the *x*–*z* plane to remove the points corresponding to false objects. Specifically, we find the largest connected component as ***ObjRegion*** for points within each candidate region list. We then find the minimum bounding box to cover the ***ObjRegion***.

For the traffic sign class, we find the convex hull around the points in the corresponding ***ObjRegion***. We then compute the ratio of the area of the minimum bounding box to the area of the convex hull to decide whether the traffic sign class corresponds to the false objects. If the ratio is larger than a predefined threshold (e.g., 0.7), it indicates that the two areas are similar; therefore, we consider ***ObjRegion*** as a traffic sign. Otherwise, we consider ***ObjRegion*** as the false object and remove it from the final results.

For the light pole class, we calculate the ratio of the length to the width of the minimum bounding box. If this ratio is larger than a predefined threshold (e.g., 5), it indicates ***ObjRegion*** is a tall rectangle; therefore, we consider it as a light pole. Otherwise, we consider ***ObjRegion*** as the false object and remove it from the final results.

4.2 Data Evaluation

4.2.1 Overview

We evaluate the performance of the proposed traffic sign and light pole detection method by conducting various experiments on eight LiDAR data sets corresponding to eight miles of the I-15 highway. We denote the name of each data set by a three-digit number followed by a hyphen and another three-digit number. The first three-digit number shows the starting milepost of the I-15 highway; the second three-digit number shows the ending milepost. The names of the data set in our experiments are *304-305*, *305-306*, *258-259*, *259-260*, *261-262*, *263-264*, *247-*

248, and *262-263*, respectively. The ground truth of the traffic signs and light poles for each data set is provided by manual inspection.

4.2.2 Qualitative Evaluation

In this section, we present the traffic sign and light pole extraction qualitative results for eight representative sections of the data sets *304-305* and *305-306* before and after involving the postprocessing step in Figures 4.1 and 4.2. Furthermore, we provide the qualitative results of the other six data sets after the postprocessing step in Figure 4.3. For all experiments, as discussed in Section 4.1, we divide the data set into 24 sections and separately process each section and concatenate the results of the sections to obtain the desired output.

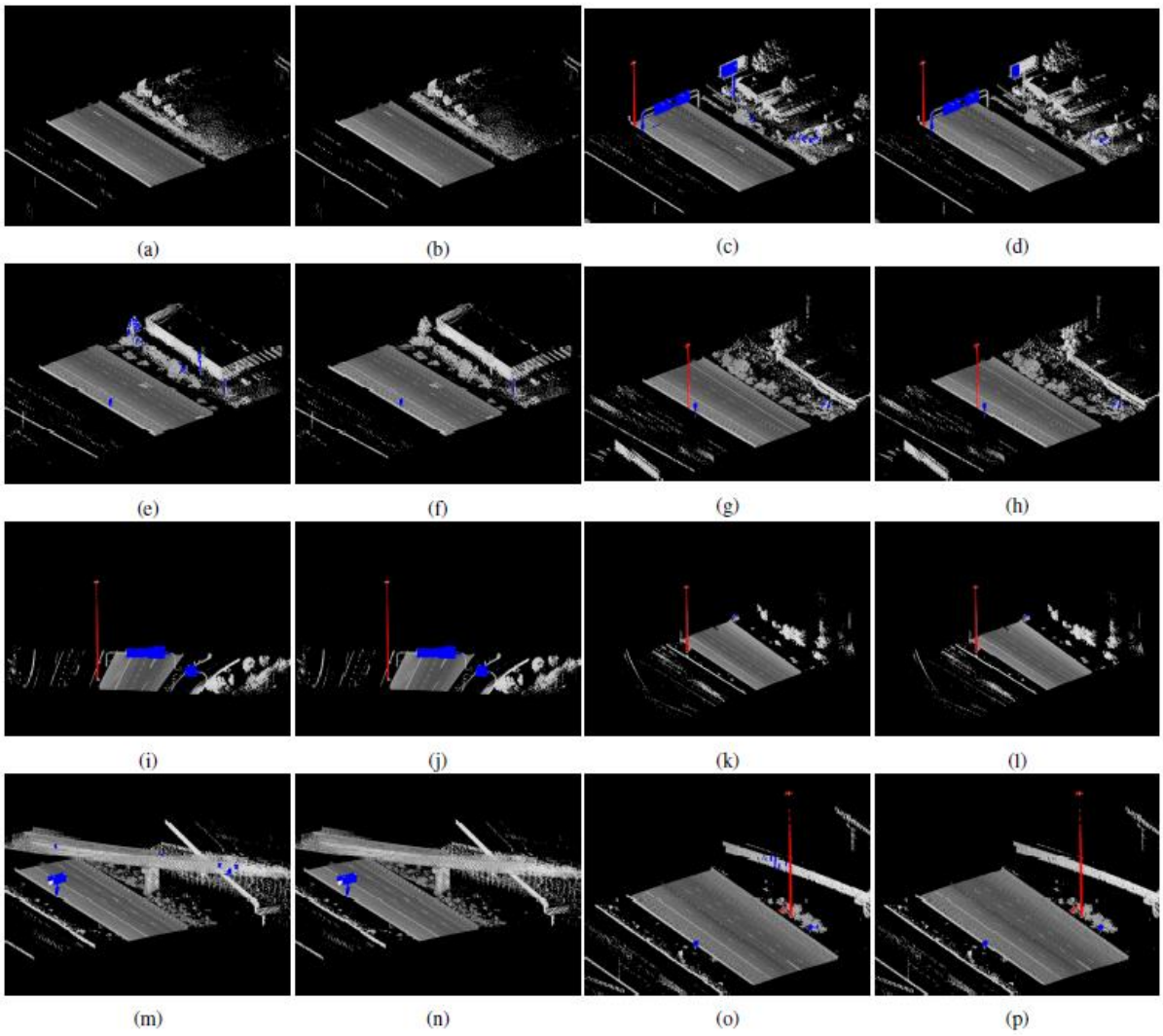

Figure 4.1 Traffic sign and light pole extraction results of the proposed method before postprocessing vs. after postprocessing for eight representative sections of the data set *304-305*. (a) vs. (b) for section 2; (c) vs. (d) for section 3; (e) vs. (f) for section 5; (g) vs. (h) for section 8; (i) vs. (j) for section 13; (k) vs. (l) for section 15; (m) vs. (n) for section 19; and (o) vs. (p) for section 22. Extracted traffic signs are shown in blue; extracted light poles are shown in red.

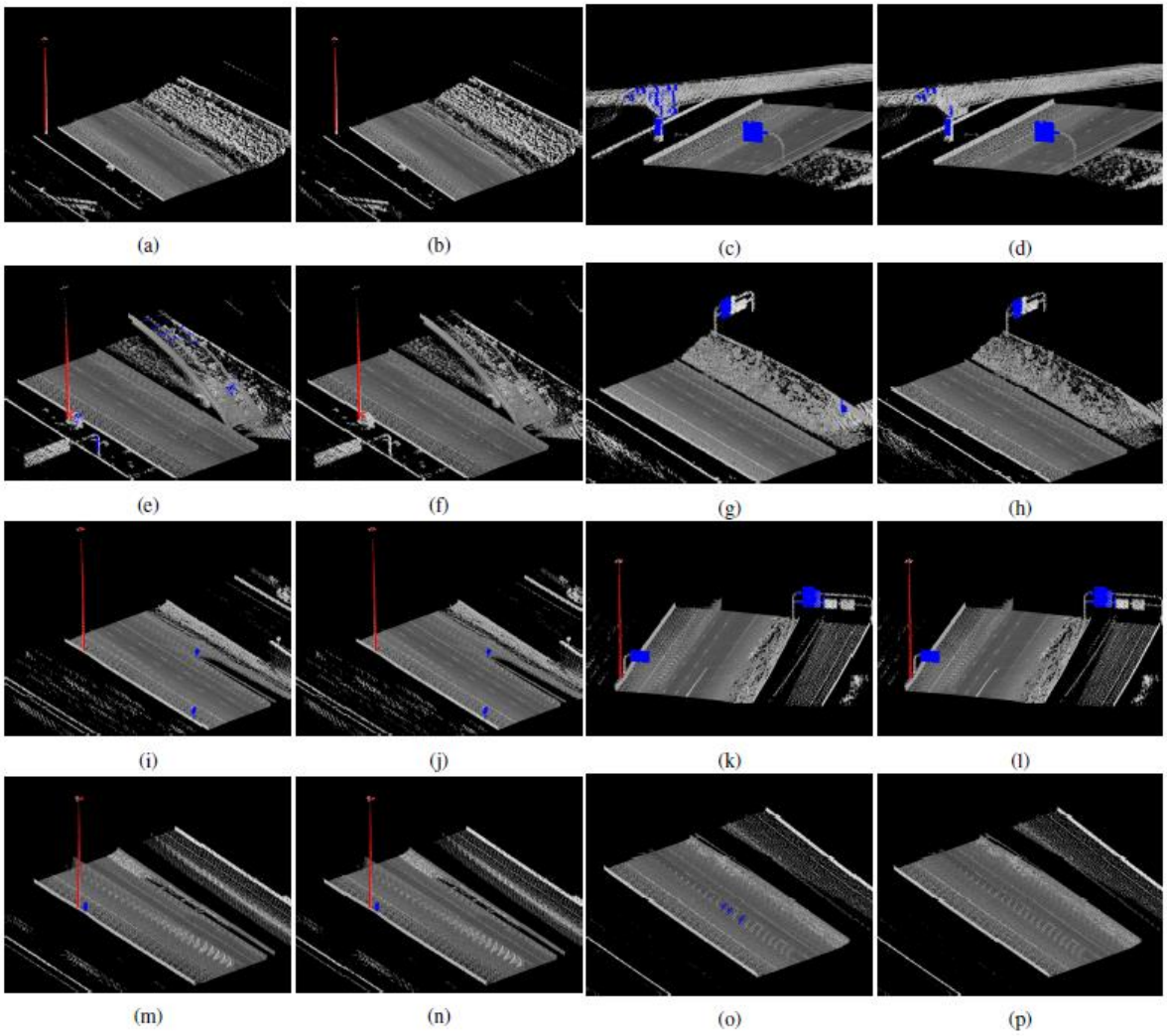

Figure 4.2 Traffic sign and light pole extraction results of the proposed method before postprocessing vs. after postprocessing for eight representative sections of the data set *305-306*. (a) vs. (b) for section 3; (c) vs. (d) for section 6; (e) vs. (f) for section 7; (g) vs. (h) for section 8; (i) vs. (j) for section 12; (k) vs. (l) for section 15; (m) vs. (n) for section 19; and (o) vs. (p) for section 21. Extracted traffic signs are shown in blue; extracted light poles are shown in red.

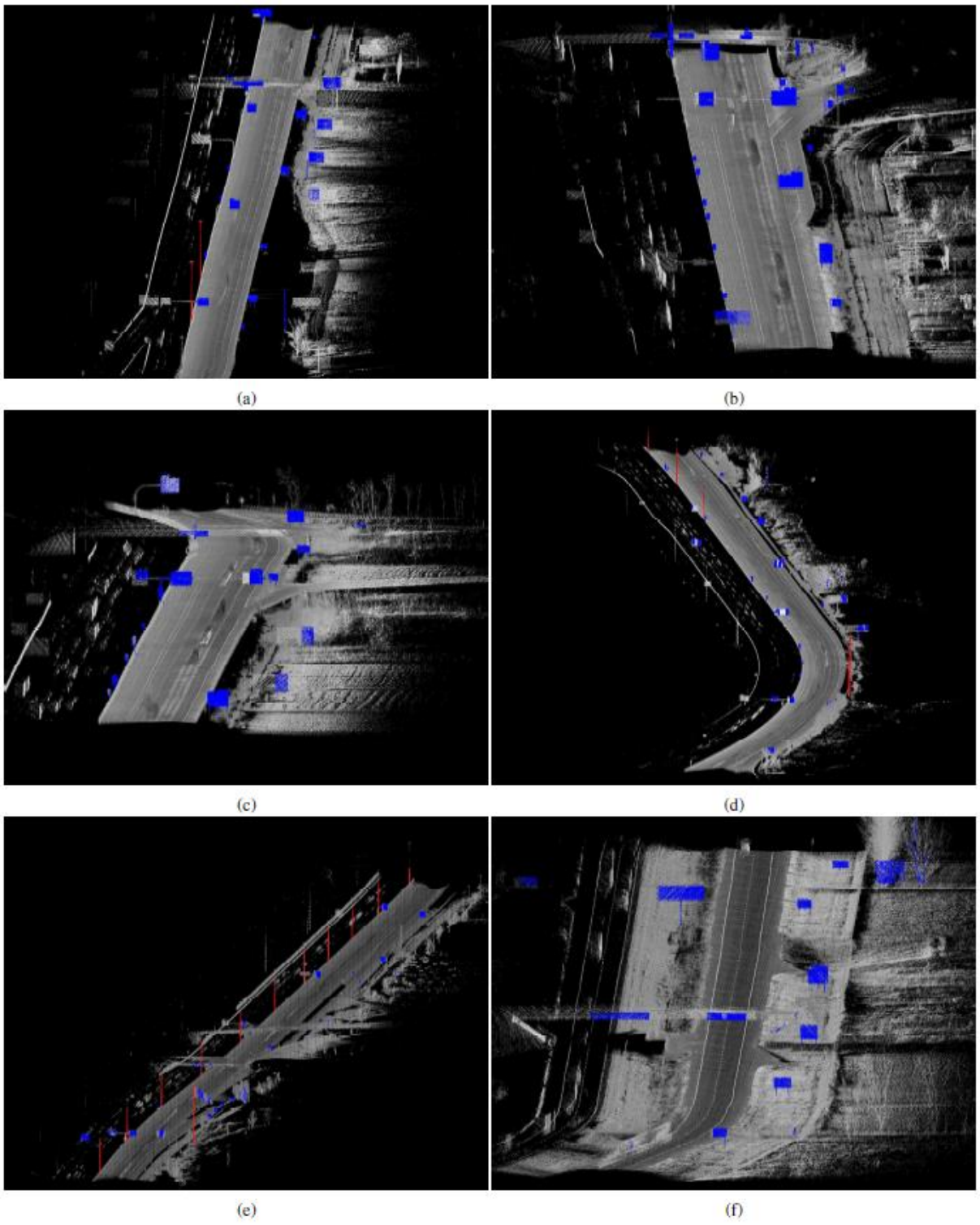

Figure 4.3 Illustration of final (after postprocessing step) traffic sign extraction results in blue and light pole extraction results in red on six data sets. (a) *258-259*; (b) *259-260*; (c) *261-262*; (d) *262-263*; (e) *263-264*; and (f) *247-248*.

4.2.3 Quantitative Evaluation

We compare the extracted traffic signs and light poles with the ground truth of their counterparts to quantitatively evaluate the performance of the proposed method and its variant method. Table 4.1 lists the ground truth and the true and false positives of the proposed method before and after the postprocessing step on all eight data sets. It clearly shows that the proposed method with the postprocessing step successfully extracts 137 (e.g., 94.48%) traffic signs and 33 (e.g., 89.19%) light poles, and the proposed method without the postprocessing step successfully extracts 139 (e.g., 95.86%) traffic signs and 33 (e.g., 89.19%) light poles. The proposed method with the postprocessing step incorrectly extracts 29 false objects, while the proposed method without the postprocessing step incorrectly extracts 70 false objects. It is clear that the proposed method detects most of the traffic signs and light poles prosperously. Moreover, it significantly reduces the number of false objects due to the use of the postprocessing step.

Table 4.1 True and false positives of the proposed traffic sign and light pole extraction method (after postprocessing step) and its variant method (before postprocessing step) on all eight data sets.

Dataset	Ground Truth		Variant Method			Proposed Method		
	Traffic Sign	Light Pole	Traffic Sign	Light Pole	False Object	Traffic Sign	Light Pole	False Object
258-259	16	3	16	2	4	16	2	2
259-260	20	0	19	0	3	19	0	1
261-262	20	0	18	0	6	18	0	3
262-263	22	5	22	4	6	22	4	3
263-264	19	12	18	11	12	18	11	5
247-248	7	0	7	0	7	7	0	3
304-305	23	8	22	8	25	21	8	9
305-306	18	9	17	8	7	16	8	3
Total	145	37	139	33	70	137	33	29

To further evaluate the proposed method, we provide the confusion matrix along with six evaluation measures, including the number of missed objects, the number of false objects, recall (detection rate), precision, *quality*, and F_1*-Measure* in Table 4.2. This confusion matrix shows that the proposed method (after postprocessing step) correctly classifies traffic signs and light poles in their corresponding classes. However, some false objects are predicted as traffic signs or

light poles mistakenly. The proposed method also misses a few objects (i.e., eight traffic signs and four light poles). The average recall, precision, *quality*, and F_1*-Measure* of traffic sign and light pole extraction results are 91.84%, 87.85%, 81.31%, and 89.68%, respectively.

Table 4.2 Confusion matrix along with six evaluation measures for each of the two classes in all eight data sets.

		Predicted	
		Traffic Sign	Light Pole
Actual	Traffic Sign	137	0
	Light Pole	0	33
Missed Objects		8	4
False Objects		26	3
Recall		94.48%	89.19%
Precision		84.04%	91.67%
Quality		80.11%	82.50%
F_1*-Measure*		88.95%	90.41%

5.0 TRAFFIC SIGN DETECTION USING REGISTERED GEO-REFERENCED RGB IMAGES AND AIRBORNE LIDAR DATA

5.1 Framework of the Proposed Method

In this project, we design and implement a method that utilizes the complementary information captured from airborne RGB images and airborne LiDAR data to detect traffic signs in highway areas. This method consists of three major components: road extraction, traffic sign candidate detection, and traffic sign classification.

- **Road Extraction:** Extract the road in the following three steps:
    - Segment the airborne images to road and nonroad regions by integrating various types of local features in a quadratic optimization model with inequality constraints.
    - Learn the road regions using sparse dictionary learning and employ morphological operations to refine the segmentation results.
    - Map the candidate road regions in airborne images to their counterparts in the corresponding airborne LiDAR data using the image to global coordinate projection.

- **Traffic Sign Candidate Detection:** Detect traffic sign candidates in the following three steps:
    - Incorporate the height information of the road points to extract the high elevated objects above the road regions.
    - Segment the extracted high elevated objects to traffic sign candidates using the Euclidean distance-based clustering algorithm.
    - Map the traffic sign candidates with enough number of points to their counterparts in the airborne RGB images.

- **Traffic Sign Classification:** Classify the candidates in the following two steps:
    - Use the VGG19 pretrained convolutional neural network (CNN) to extract local deep features to represent traffic sign candidates.

- Incorporate the local deep features in a designed local embedded convex optimization framework to classify each representation as one of the two classes: the traffic sign class and the nontraffic sign class.

In the following, we explain each of the three components of the proposed method. We illustrate the experimental results at each of the important stages on a sample pair of airborne RGB images and airborne LiDAR data.

5.1.1 Road Extraction

The size of the input geo-referenced RGB image is too large to be fed into any segmentation algorithm. As a result, we select a region of the input image that contains the road and resize it to 25% of its actual size as a necessary preprocessing step. Different complementary local features, including color, gradients, soft segmentation, and texton, are individually extracted and then combined to represent each region. These combined local features are represented in a quadratic optimization problem with inequality constraints. Using the joint feature representation, segmentation results can be easily obtained by labeling each region to the index of the largest element of the corresponding column. To refine the segmentation results, we use the two prebuilt dictionaries, which are, respectively, constructed by a sufficient number of road points and nonroad points, to keep the candidate road regions that are well-represented by the prebuilt road dictionary (i.e., the regions with the low reconstruction error). We further apply morphological operations to find the road region. To achieve this goal, we find the connected components and keep the largest connected component as the final main road section result because the main road covers the larger area of the input image. We map the candidate road sections in airborne images to their counterparts in the corresponding LiDAR data using the image to the global coordinate projection. To this end, we use affine transformation parameters to construct the transformation matrix between pixel indices in the airborne RGB images and the global coordinates. We then convert the global coordinate to the latitude/longitude coordinate system of the airborne LiDAR data. Figure 5.1 shows the input airborne image (top left), its input airborne LiDAR data (top right), the segmentation results in the airborne image (bottom left), the road section result in the airborne image after employing the road point sparse

dictionary learning and morphological refinement (bottom middle), and the extracted road region in the LiDAR data after projection (bottom right).

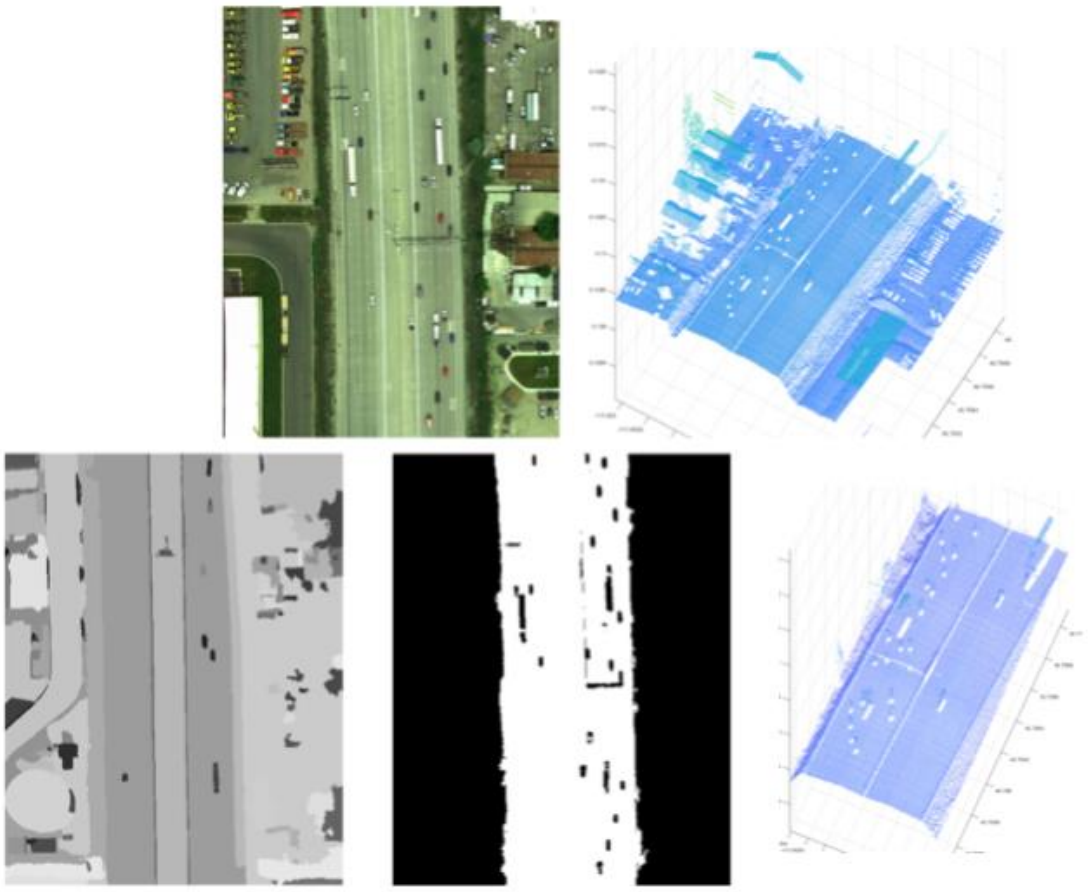

Figure 5.1 Sample pair of airborne images (selected from *i15-north-rect-r1-c1* mosaic) and its corresponding LiDAR data (top row) and their sample road extraction results (bottom row).

5.1.2 Traffic Sign Candidate Detection

We utilize the height (z) information of the extracted road points in the LiDAR data to extract the high elevated objects above the road sections. To achieve this, we first calculate the histogram of the z value of the extracted road points in the LiDAR data. We then find the center of the bin of the histogram, which has the maximum number of points, and set the center value as the threshold T_z. Next, we filter out the points whose z values are less than $T_z + T_0$, where T_0 is empirically determined to filter out the cars or any low elevated objects on the road. Finally, we use the Euclidean distance-based clustering algorithm to segment the extracted high elevated objects to traffic sign candidates. We further remove the segments containing less than 50 points because the traffic sign regions are usually in a sizeable rectangular shape. We finally map the outlier removed traffic sign candidates to the airborne RGB images using the LiDAR to image pixel indices projection. Figure 5.2 demonstrates two of the extracted traffic sign candidates in the airborne image for the selected region.

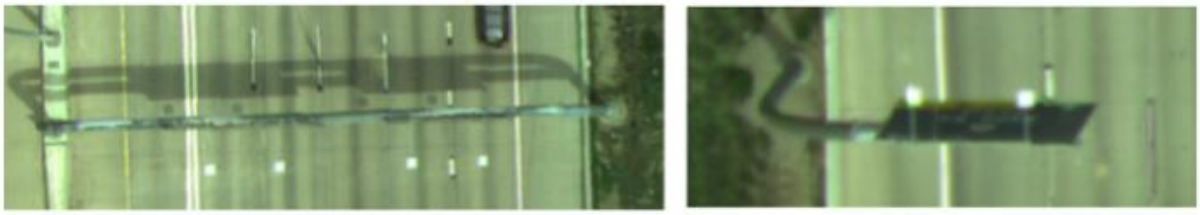

Figure 5.2 Two traffic sign candidates extracted in the selected region of the sample airborne image.

5.1.3 Traffic Sign Classification

Due to the small number of available images in our data set, training a CNN for the classification task cannot lead to an accurate model. Instead, we employ a pretrained deep neural network, which is trained on more than 1 million images categorized into 1,000 classes. We use this pretrained network to extract the deep features for each traffic sign candidate. To this end, we resize the traffic sign candidates to 112×112×3 and pass them to the VGG-Net 19. We extract the features from layer *Conv 5-4* and resample the output to the size of 28×28×512. We further

construct nine local feature maps with the size of 14×14×512 using the stride of 7 from the resampled output. We then flatten each local feature map into a vector and perform PCA to obtain the top 1,120 features. We utilize the deep local features in a designed local-embedded convex optimization problem to represent the candidates by a set of traffic sign templates, which are constructed by randomly selecting 15 traffic signs in the data set. A pooling strategy is finally employed to assign a score to each candidate. A candidate with a higher score than a predefined threshold is considered as a traffic sign. The two traffic sign candidates shown in Figure 5.2 are classified as belonging to the traffic sign class because their scores are greater than the predefined threshold.

5.2 Data Evaluation

5.2.1 Overview

In this section, we evaluate the performance of the proposed method on nine different sections of the I-15 highway by providing qualitative and quantitative results, including true positives, false positives, and true negatives.

5.2.2 Qualitative Evaluation

In this section, we provide road and traffic sign extraction results on eight additional sections of multiple pairs of airborne images and LiDAR data. Specifically, we present the sample pair of airborne image and LiDAR data (top row), their road extraction results (middle row), and the traffic sign candidate extraction results in the airborne image (bottom row) in Figures 5.3 through 5.9, respectively. We also label the classification result of each traffic sign candidate at the bottom left of each airborne image in the bottom row of each figure, with "TS" indicating the traffic sign and "NTS" indicating the nontraffic sign.

5.2.3 Quantitative Evaluation

In addition, we provide the true positives, false positives, and true negatives of the proposed method on all the nine sections of the data set. In total, there are 17 traffic signs in these sections. The proposed method extracts 24 traffic sign candidates. In the classification

process, 14 out of 24 candidates are correctly classified as traffic signs (true positives), three out of 24 candidates are incorrectly classified as traffic signs (false positives), and seven out of 24 candidates are correctly classified as nontraffic signs (true negatives). In other words, the proposed method is able to successfully extract 14 out of 17 traffic signs and achieve the detection accuracy of 82.35%.

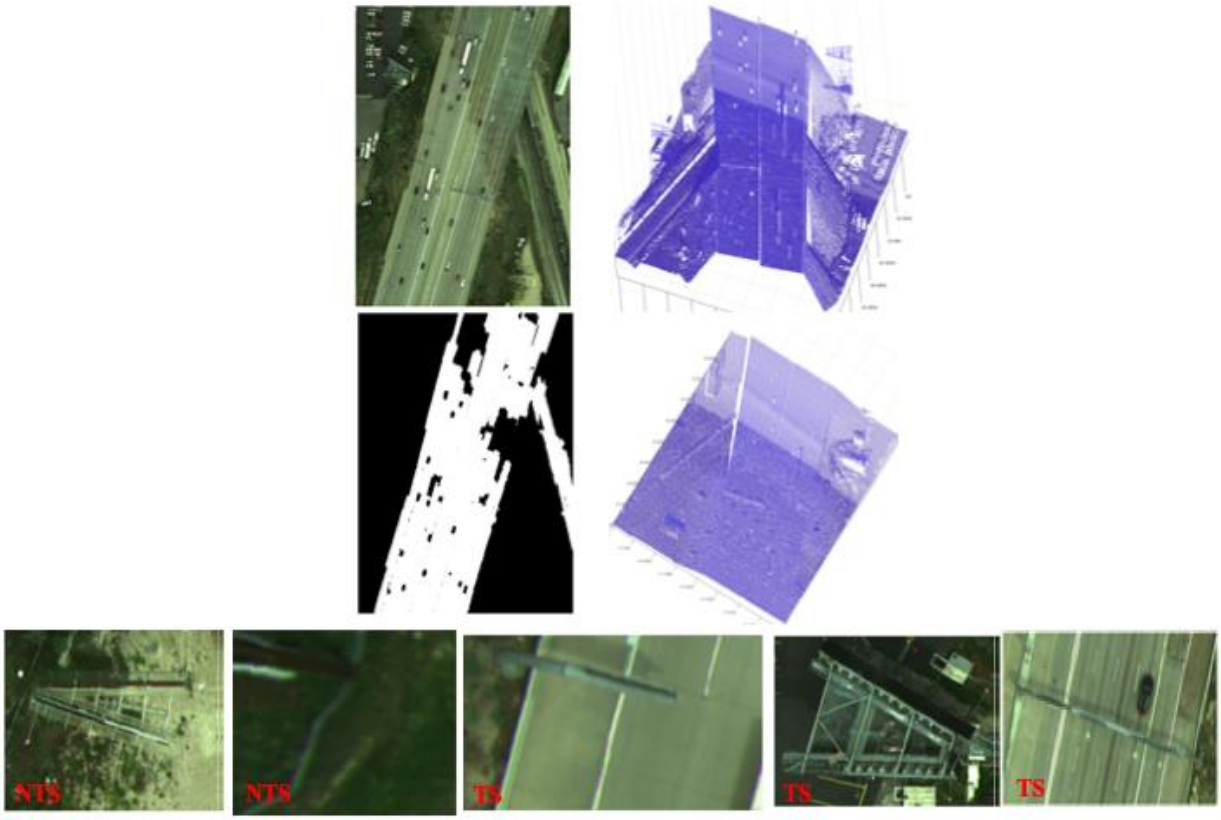

Figure 5.3 *Section2* (selected from *i15-north-rect-r2-c1* mosaic): Airborne image and its corresponding LiDAR data (top row) together with their road extraction results (middle row) and the traffic sign candidate extraction results in the airborne image (bottom row). Two traffic signs are present in this data set. The proposed algorithm extracts five traffic sign candidates and correctly classifies two of them as traffic signs (labeled as TS in the left bottom) and two of them as nontraffic signs (labeled as NTS in the left bottom). It incorrectly classifies one of the candidates as traffic sign (labeled as TS in the left bottom).

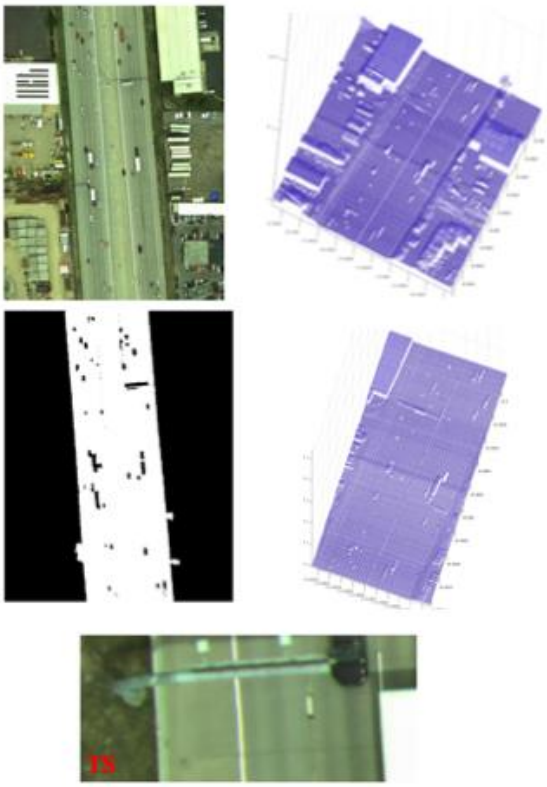

Figure 5.4 *Section3* (selected from *i15-north-rect-r3-c1* mosaic): Airborne image and its corresponding LiDAR data (top row) together with their road extraction results (middle row) and the traffic sign candidate extraction results in the airborne image (bottom row). Two traffic signs are present in this data set. The proposed algorithm extracts one traffic sign candidate and correctly classifies it as traffic sign (labeled as TS in the left bottom).

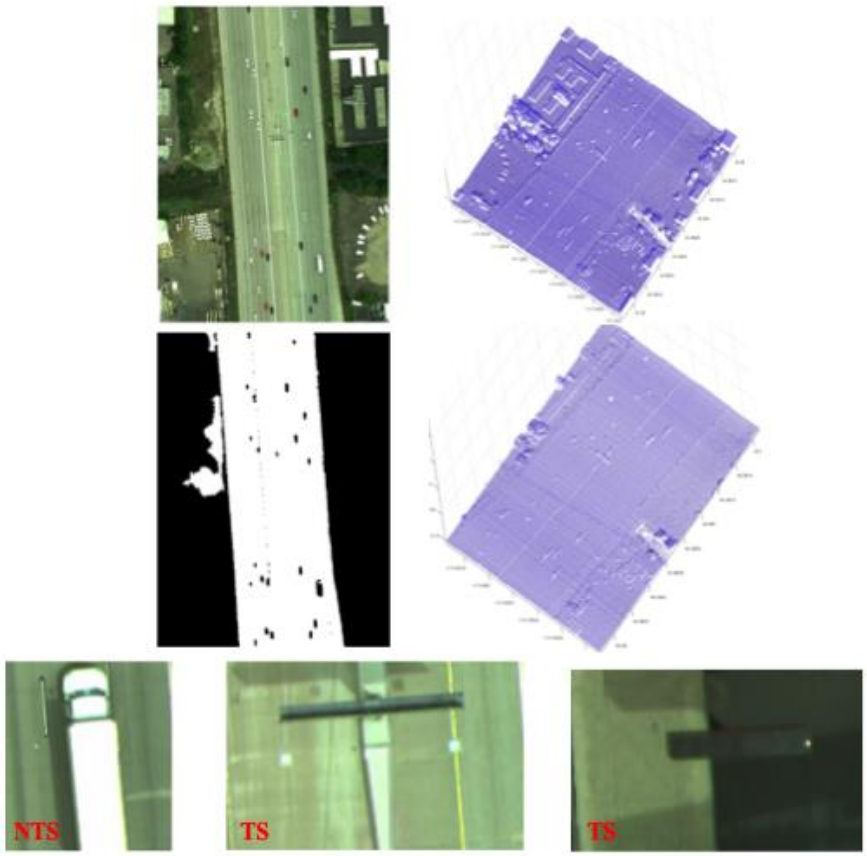

Figure 5.5 ***Section4*** **(selected from** ***i15-north-rect-r3-c1*** **mosaic): Airborne image and its corresponding LiDAR data (the top row) together with their road extraction results (middle row) and the traffic sign candidate extraction results in the airborne image (bottom row). One traffic sign is present in this data set. The proposed algorithm extracts three traffic sign candidates and correctly classifies one of them as the traffic sign (labeled as TS in the left bottom) and one of them as the nontraffic sign (labeled as NTS in the left bottom). It incorrectly classifies one of the candidates as the traffic sign (labeled as TS in the left bottom).**

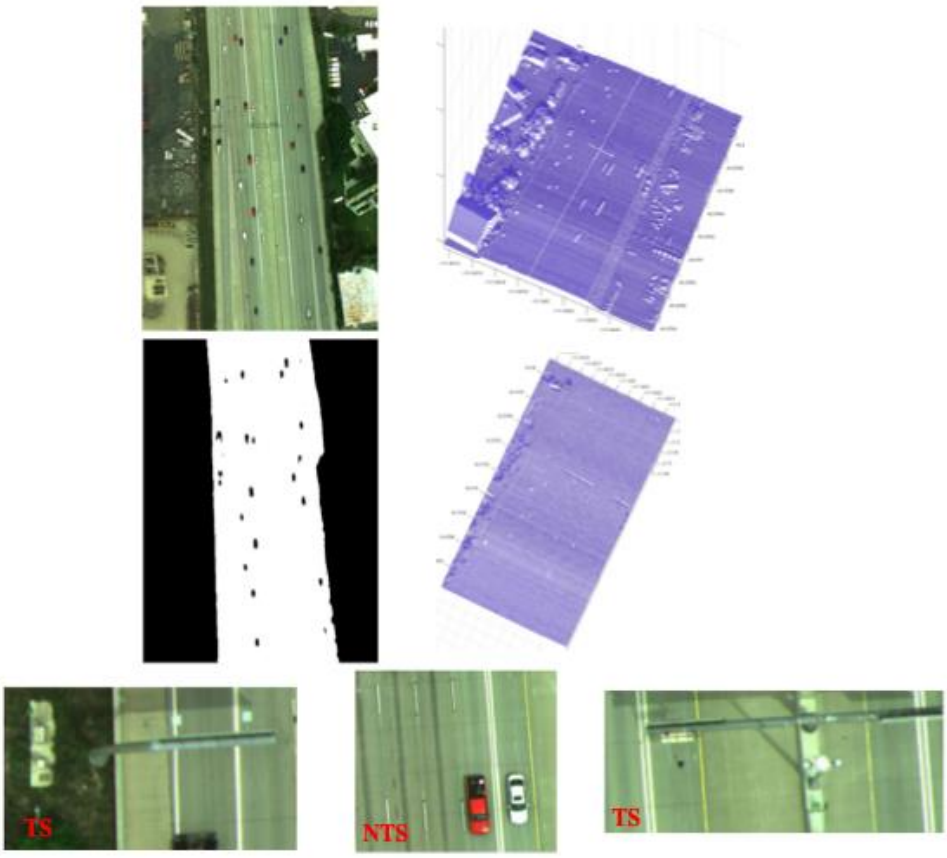

Figure 5.6 ***Section 5*** **(selected from** ***i15-north-rect-r3-c1*** **mosaic): Airborne image and its corresponding LiDAR data (the top row) together with their road extraction results (middle row) and the traffic sign candidate extraction results in the airborne image (bottom row). Two traffic signs are present in this data set. The proposed algorithm extracts three traffic sign candidates and correctly classifies two of them as traffic signs (labeled as TS in the left bottom) and the other one as the nontraffic sign (labeled as NTS in the left bottom).**

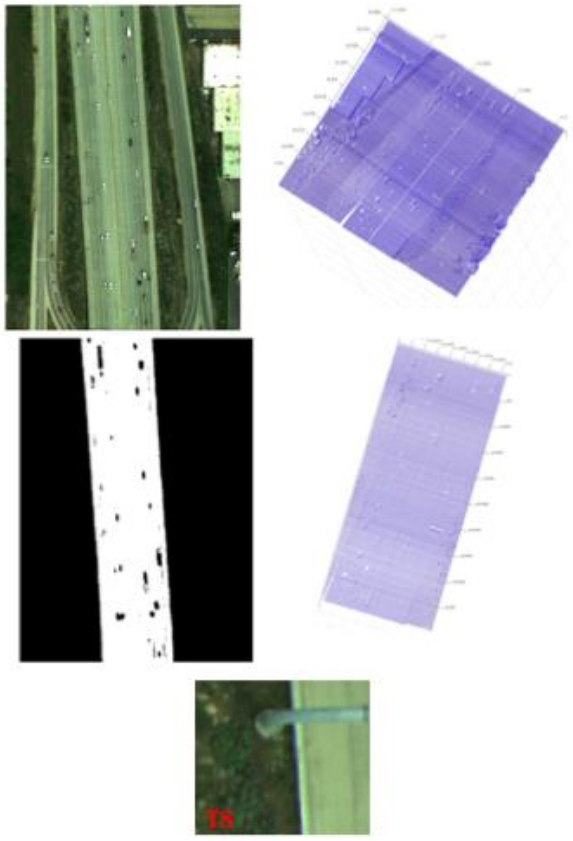

Figure 5.7 *Section6* (selected from *i15-north-rect-r3-c1* mosaic): Airborne image and its corresponding LiDAR data (top row) together with their road extraction results (middle row) and the traffic sign candidate extraction results in the airborne image (bottom row). One traffic sign is present in this data set. The proposed algorithm extracts one traffic sign candidate and correctly classifies it as the traffic sign (labeled as TS in the left bottom).

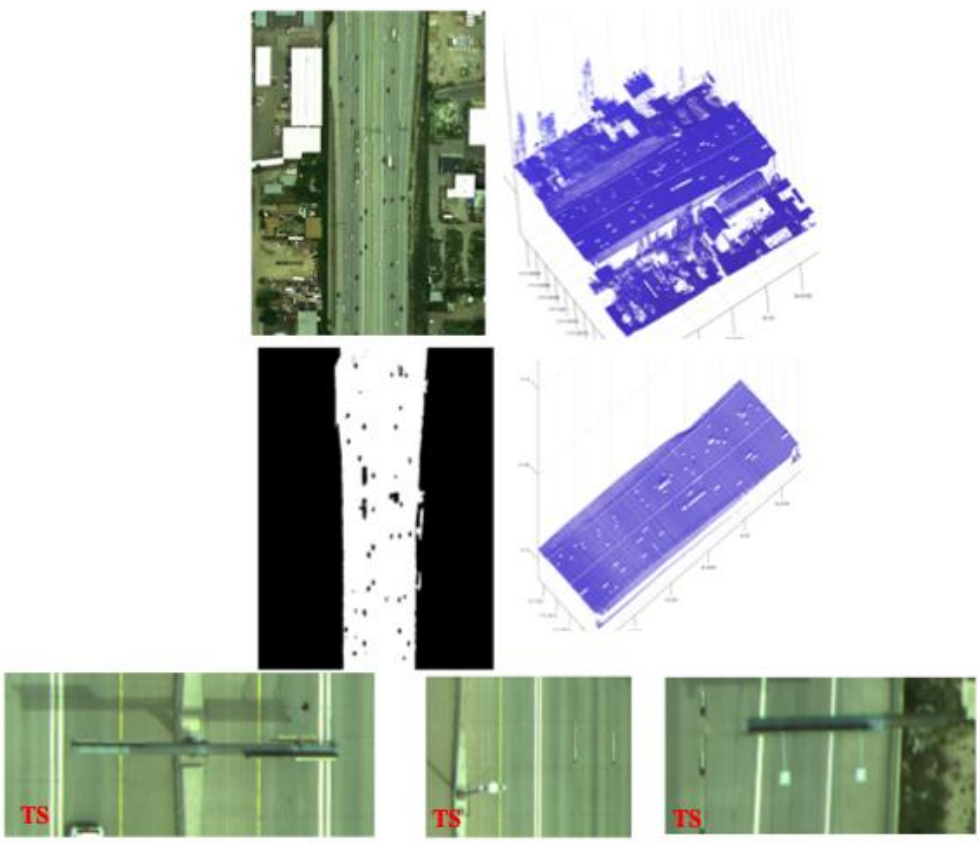

Figure 5.8 ***Section7*** **(selected from** ***i15-north-rect-r4-c1*** **mosaic): Airborne image and its corresponding LiDAR data (the top row) together with their road extraction results (middle row) and the traffic sign candidate extraction results in the airborne image (bottom row). Three traffic signs are present in this data set. The proposed algorithm extracts three traffic sign candidates and correctly classifies two of them as traffic signs (labeled as TS in the left bottom). It incorrectly classifies one of candidates as the traffic sign (labeled as TS in the left bottom).**

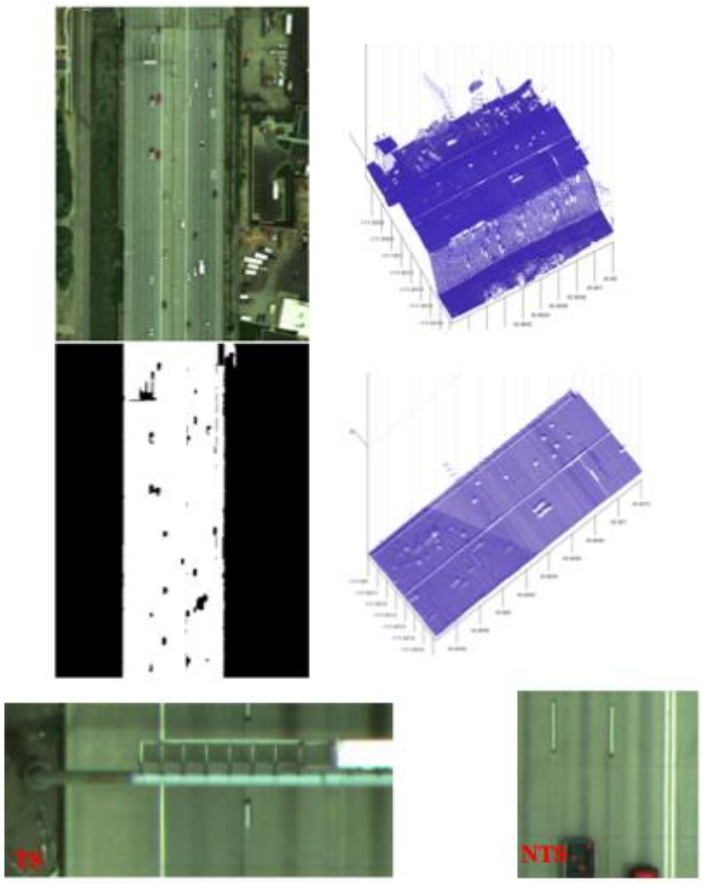

Figure 5.9 ***Section 8*** **(selected from** ***i15-north-rect-r4-c1*** **mosaic): Airborne image and its corresponding LiDAR data (top row) together with their road extraction results (middle row) and the traffic sign candidate extraction results in the airborne image (bottom row). Two traffic signs are present in this data set. The proposed algorithm extracts two traffic sign candidates and correctly classifies one of them as the traffic sign (labeled as TS in the left bottom) and the other one as the non-traffic sign (labeled as NTS in the left bottom).**

6.0 CONCLUSION

6.1 Overview

In this section, we present the conclusion of this research, important findings, and the most important challenges during the project.

6.2 Conclusion

In the first stage of this project, we propose a fast and reliable traffic sign and light pole detection method, which can be applied to the MLS data to quickly identify various traffic signs and light poles. A set of experiments has been carried out on the eight data sets that are captured by USU along the I-15 highway. The extensive experimental results demonstrate that the proposed method is able to successfully detect 137 (e.g., 94.48%) traffic signs and 33 (e.g., 89.19%) light poles in the eight data sets. In other words, the proposed method is robust in detecting almost all traffic signs and light poles with a few numbers of false positives. Our contributions are: 1) employing the surface reconstruction algorithm to extract the orientation of the points as one of the characteristic features; 2) applying the unsupervised *k*-means clustering algorithm to automatically extract road points; 3) designing a sliding cuboid to search for the high elevated objects above or beside the roads as groups of candidate points; 4) employing the RANSAC algorithm to select the robust candidate points that represent planes along the vehicle trajectory; 5) proposing a modified seeded region growing algorithm to remove the outlier points around the objects; 6) introducing a shaped-based false object rejection algorithm to remove the false positive objects.

In the second stage of this project, we effectively fuse the complementary information of airborne geo-referenced images and airborne LiDAR data to accurately detect the traffic signs along the I-15 highway in the state of Utah. Our designed method consists of three main components: 1) road extraction; 2) traffic sign candidate detection; 3) traffic sign classification. Two uniquely designed convex optimization models are, respectively, employed in the road extraction and traffic sign classification components to accurately identify road candidates and traffic signs. The joint local features and the deep local features are also seamlessly employed in

these two optimization models to provide a more accurate representation of the road and the traffic signs, respectively. The experimental results demonstrate that the proposed framework performs well in detecting traffic signs.

6.3 Findings

Some of the important findings in this research project are summarized as follows:

- Using the complementary information form RGB images and airborne LiDAR data helps to more accurately detect traffic signs in the I-15 highway areas.
- Extracting the road regions is an essential initial step, which helps to reduce the searching areas for the traffic signs.
- Representing the deep local features in a designed local-embedded convex optimization problem helps to classify each traffic sign candidate as either traffic sign class or nontraffic sign class.

6.4 Challenges

The most important challenges of this research work are summarized below:

- Registration of the RGB images and the airborne LiDAR data, which is necessary for any further processing.
- Design and implementation of an image-processing technique to automatically extract road regions from the mosaic maps.
- Similarity between the road regions and other regions in the RGB images in terms of color and texture.
- Low density in airborne LiDAR data sets for objects such as traffic signs and buildings due to the high altitude of the airplane used to capture the LiDAR data.
- Large number of outlier points in the LiDAR data set.
- Design of an optimization problem to model the deep local features of traffic signs.
- Solving the optimization problem numerically, which is essential for correct classification of traffic sign candidates.

In summary, we investigated the potential of a data fusion approach of combining the complementary features acquired from both on-board cameras and LiDAR sensors for extracting highway maintenance features. The results demonstrated the robustness of the proposed data fusion approach. A number of research extensions can be considered in future studies. First, with the advent of UAV-based LiDAR technology, we would like to test the proposed data fusion approach on an UAV-based LiDAR platform. Second, we plan to test more powerful deep learning techniques to improve the accuracy of our algorithms.

REFERENCES

[1] He, Y., Song, Z., Liu, Z., and Lindsey, R. (2016). *Implementation of aerial LiDAR technology to update highway feature inventory* (No. Report No. UT-17.06). Utah. Dept. of Transportation. Research Division.

[2] He, Y., Song, Z., & Liu, Z. (2017). Updating highway asset inventory using airborne LiDAR. *Measurement*, *104*, 132-141.

[3] He, Y., Song, Z., and Liu, Z. (2017). *Highway Asset Inventory Data Collection Using Airborne LiDAR* (No. 17-04058).

[4] Soheilian, B., Paparoditis, N., and Vallet, B. (2013). "Detection and 3D Reconstruction of Traffic Signs from Multiple View Color Images." *ISPRS Journal of Photogrammetry and Remote Sensing*, no. 77, 1–20.

[5] Adam, A. and Ioannidis, C. (2014). "Automatic Road Sign Detection and Classification Based on Support Vector Machines and HOG Descriptors." *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, no. 2, vol. 5, 1-7.

[6] Khalid, S., Muhammad, N., and Sharif, M. (2019). "Automatic Measurement of the Traffic Sign with Digital Segmentation and Recognition." *IET Intelligent Transport Systems*, vol. 13, no. 2, 269-279.

[7] Pu, S., Rutzinger, M., Vosselman, G., and Elberink, S.O. (2011).: "Recognizing Basic Structures from Mobile Lasers Scanning Data for Road Inventory Studies." *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 66, no. 6, S28–S39.

[8] Yokoyama, H., Date, H., Kanai, S., and Takeda, H. (2013). "Detection and Classification of Pole-Like Objects from Mobile Laser Scanning Data of Urban Environments." *International Journal of CAD/CAM*, vol. 13, no. 2, 31–40.

[9] Yu, Y., Li, J., Guan, H., and Wang, C. (2015) "Automated Extraction of Urban Road Facilities Using Mobile Lasers Scanning Data." *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 4, 2167–2181.

[10] Riveiro, B., Dıaz.Vilarino, L., Conde.Carnero, B., Soil´an, and M., Arias, P. (2016). "Automatic Segmentation and Shape-Based Classification of Retro-Reflective Traffic Signs from Mobile LiDAR Data." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 9, no. 1, 295–303.

[11] Lehtomaki, M., Jaakkola, A., Hyypp¨a, J., Lampinen, J., Kaartinen, H., Kukko, A., Puttonen, E., and Hyyppa, H. (2016). "Object Classification and Recognition from Mobile Laser Scanning Point Clouds in a Road Environment." *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 2, 226–1239.